

《心理学报》审稿意见与作者回应

题目：基金投资决策中老手与新手的加工差异及干预
作者：辛自强; 王鲁晓; 李越

第一轮

审稿人 1 意见：

本研究采用实验室实验法，讨论基金投资决策的信息加工特征，并通过“专家-新手”比较范式揭示其基金投资决策结果的差异以及背后信息加工方式的差异。研究问题有一定的意义，但该论文具有比较明显的问题，那就是研究变量的设计有多处不合理的地方，导致研究结果存在较多疑问。

意见 1：从自变量“基金属性”来看，作者对此处实验材料的设计比较粗糙，然而这是本研究的关键。基金是非常复杂的，不同 APP 采用的评级体系或标准不尽相同，作者选用的是什么 APP？是否可靠？参考了什么评价体系？作者需详细介绍，并且论证所选取的属性关键可靠，仅依靠被试去选择出关键属性还不足以佐证实验材料设计的可靠性。

回应：

首先感谢您的鼓励和所提出的问题。这里您提到基金属性的获取方式以及选取是否可靠这两个问题，这也是我们在设计实验材料时主要考虑的内容。我们之前在论文中表述不够清晰，现已根据您的建议修改并说明修改情况：

(1)基金属性的获取途径。本研究基金属性的选取主要基于市场上常见的支付宝理财，同时参考了各大银行的基金属性信息，这一点我们在预研究中进行补充说明(2.1.1, p.5)。支付宝基金自 2013 年推出，截止 2021 年该软件的基金投资者已达 6 亿人，可见支付宝基金是投资者较为常用的理财 APP。虽然不同 APP 的基金属性及评价标准不尽相同，但大致相似。例如支付宝同时使用“近一年收益率”、“近一年涨跌幅”两种表述对某基金在过去一年的总收益情况进行说明，其他渠道(如微信理财通、招商银行基金等)也同样采用“近一年涨跌幅”进行描述。

(2)选取方式是否可靠。我们通过调查基金新手和老手后，选取年收益率、风险等级、基金评级、历史业绩和日涨跌幅这 5 个基金属性，这是因为基金投资者是基金领域的重要参

与者，对基金有着较为清晰的认识，他们的选择一定程度上反映了何种属性决定了基金的优劣。此外，这些属性确实对基金投资决策十分重要，符合现实情况，图 1-1 为招商银行和支付宝基金页面截图，在基金初始界面中均可看到年收益率、风险等级、基金评级、历史业绩和日涨跌幅这 5 个重要属性。

我们推测您关心这 5 个重要属性的优劣能否反映基金质量。从长期来看，确实难以通过基金属性推测基金的收益。调查表明基金的波动率、最大回撤率以及基金经理水平等因素无法预测基金的实际投资收益(景顺长城 等, 2021)，但这不是本研究的主题。我们致力于探讨投资者进场的问题，即关注个体在最初决定投资基金时的推理过程，在本研究的实验任务中则表现为存在理论上的最优选项的情况下，投资者能否选中最优选项及其背后的认知推理过程。因此 5 个基金属性的设置是为了提供逻辑推理的标准，确实难以完全预测基金的长期质量，这一点我们在局限中进行了说明(4.4 第一段, p.17)。



图 1-1 招商银行(左)和支付宝(右)基金页面

意见 2：对于因变量“决策质量”，研究中的定义看起来也很模糊。作者是直接选取的 APP 的评级还是自己根据属性设定来评的级？APP 的评级并未只基于这五个属性，作者设定的评级与其他属性是否能匹配？另外，五星和四星、三星的差异有多大？为何只有五星标记为 1？

回应：

感谢您提的问题。基金评级与其他四个属性并列，均通过人为模拟的方式确定数值，并非根据其他属性计算得来。为了减少歧义并降低阅读难度，我们对 2.4.1 第二段 (p.6)的措辞进行了斟酌、调整。

5 个属性并列是基于现实情况。一方面基金评级与诸多因素(如基金类型)有关，并非完全依赖于本研究所选择的涨跌幅、风险程度等基金属性，但我们进行实验模拟研究时只能选取有限的属性设定任务；另一方面，投资者会综合若干属性信息后进行决策，而非主要依赖基金评级，正如预研究发现仅有 36.2%的投资者认为基金评级十分重要。从现实和个人心理上，基金评级都非其他属性的上位概念，因此在模拟时无需将基金评级与其他属性进行匹配。

为更直观地探究决策质量，我们在基金投资任务中人为设置了最优选项，基金 443 在各个基金属性上均表现最优，其他选择则各有劣势。在确定属性等级时，我们同样参考了支付宝属性评级的呈现方式，即收益率和涨跌幅通过百分比呈现，其余属性通过等级呈现(相关内容补充在 2.4.1, p.6)。每个属性的优劣均可通过对比得来，如四星高于三星。

意见 3: 对于指标 SM 值，对于本研究来说是否合理？有无以往相似研究这样运用。在本研究中 SM 的使用可能存在偏差，因为从实验设计来看，似乎很难让被试基于选项进行转换，因为选项是一个无意义的代码，被试只能根据属性去选择点开感兴趣的框，竖着点开就是基于维度？横着点开就是基于选项吗？作者应进行更清晰的阐述。

回应:

感谢您的意见，之前我们并未对指标的选取标准进行明确的陈述，我们在 2.4.1 第四段中予以补充(p.7-8)。

包括 SM 值在内的所有指标都是我们阅读了相关文献后选取的。就 SM 值而言，这是决策研究领域过程追踪的重要指标(余雯 等, 2013)，用以反映被试在决策过程中是基于选项还是基于属性进行信息搜索，以此作为启发式决策策略的检验(Schulte-Mecklenbeck et al., 2011)，投资者在基金决策过程中的信息加工策略正是本研究的主题。以往研究者也采用此方法和指标对复杂决策过程进行探究，例如国内研究者王阿妹等人(2018)将应聘者作为选项，执行力、责任感等特征作为属性探究被试招聘决策中的搜索模式，冷静等人(2017)将职业作为选项，距离、工资等特征作为属性探究被试在求职决策过程中的搜索模式，这都与本研究的逻辑是一致的。

感谢您提出的“被试难以基于选项进行转换”这一问题，这是我们之前并未深入思考的。经过进一步文献检索，我们认为当前的呈现方法并不意味着被试难以基于选项进行转换，有

三方面原因：一为以往研究多采用这种呈现方法，是相对可靠的；二为之前也出现过被试基于选项进行加工的结果；三为基于选项加工才符合人们的横向阅读习惯(Tullis, 1988)。我们在讨论中(4.1 第二段, p.15)进行了进一步阐明。同时，我们还对搜索模式的逻辑和计算进行了补充说明 (2.4.1, p.10)，计算方式具体为被试横向查看(基于选项)与纵向查看(基于属性)次数差异的函数。

意见 4: 对于研究结果和结论来看，有不少不合理之处，例如，虽然老手和新手在补偿性程度上有差异，但从结果均值来看，老手和新手的补偿性都很高。此外，本研究还存在一些写作和其他问题：例如，

引言部分，假设 3 应把新手具体表现写出来；

多处内容，例如 1.4，推理都不够严谨，需进一步细化。

又如，被试选择不严谨：(1)投资经验年限范围太窄；(2)未报告是否为基金等相关从业者；(3)区分老手和新手的问题太少，太简单，实际上，根据这些问题可能无法有效区分被试的基金投资能力；(4)有效问卷量太少，若从一个差异更广，较大的被试范围内选择前后约 30%的被试来作为老手和新手可能更合理。总之，本研究最大的问题就是变量的选取和设置存在争议，需先确保变量的准确和合理性，研究结论才有意义。

回应:

十分感谢您的意见，这启发我们在结果分析、假设推导和被试选择三方面进行深入思考。

首先，我们再次对研究 1 中新手和老手的原始数据进行深入挖掘。以搜索深度为例，虽然老手和新手组被试的搜索深度分别达到 0.90 和 0.96，但是我们发现 64%的新手会查看所有信息，而仅有 40%的老手会全部查看，这在一定程度上表明两组被试背后的补偿性确实存在差异。两组被试的搜索深度均较高可能由于实验任务使得搜索难度降低，我们在研究局限中补充说明了任务简化现实情境的情况(4.4 第一段, p.17)。

其次，我们对引言进行再次推敲，为假设推理提供更有力的证据。在 1.3 中，我们从老手的决策方式和行为表现更为稳定的角度提出假设 3，并补充了相关文献(p.3)；在 1.4 中，我们补充了表格式的信息呈现为何能够作为干预手段，并且补充了此干预手段的文献 (Spence & Brucks, 1997; p.4)。

最后，我们对被试选择进行如下解释。老手和新手的划分并无绝对标准，往往是相对的划分。以风险决策领域研究为例，Ognjanovic 等人(2019)通过是否有 3 个月及以上投行实习经验作为标准对 98 名被试进行划分，探究金融投资新手和老手的差异；国内研究者成晓君

等人(2022)将 60 名被试分为 30 组后，根据组内两人在基线任务中表现的相对好坏将其划分为专家和新手。我们通过预研究，以约 30%的老手占比进行划分，将 3 年经验作为界限也有合理性。我们主要关注经验的差异并据此划分老手和新手，其决策能力和表现是否存在差异还未可知，实际上其投资时的决策表现和结果恰恰是本文关注的重点。之前研究也表明，老手和新手的决策质量是否存在差异是存在矛盾结果的，这一点我们已在讨论中说明(见 4.1, p.15)。研究 1 中未询问被试是否为基金从业者，但是询问了被试是否为财经相关专业，以此作为财经知识的指标进行控制。

.....

审稿人 2 意见：

《基金投资决策中老手与新手的的信息加工差异及干预》一文，考察了基金投资决策中老手与新手在信息获取模式上的差异，并进行了一项干预研究。该文在把投资决策的过程追踪研究与干预研究结合的思路，有一定新意。但论文在研究逻辑、论文写作和数据分析上均存在一些问题，有待深入挖掘，并大幅度予以改进。

意见 1：中文摘要的逻辑不够清晰。可否补充以往研究中存在的问题，及其与本研究的逻辑关系？

回应：

感谢您的提醒，我们阅读了多篇近年来发表的心理学报的论文作为参考，随后我们对摘要部分进行了补充调整，指出以往研究的不足及需求，并陈述了研究 1 和研究 2 的逻辑关系(摘要, p.1)。

意见 2：论文提出的贡献是“首次使用 Mouselab 技术探讨基金投资决策中的信息加工过程特征及其与决策质量的关系”，建议作者系统回顾一下相关文献，并慎用“首次”的表述。我没有做非常系统的文献搜索，但是考虑到 MouseLab 和投资决策的研究均有一定时间了，很可能已经有过类似的研究。

同样，在综述部分，作者多次提到“缺乏有关基金投资决策过程的直接证据”、“在基金投资领域，“老手-新手”的信息加工过程差异如何体现，目前尚无直接的研究证据”。我对这些表述存疑。据我所知，用眼动比较金融投资决策及专家新手差异的研究不在少数，建议补充对相关文献的回顾。我查了两篇相关综述，列在此处，供作者参考。

Sickmann, J., & Le, N. (2016). Eye-tracking in behavioural economics and finance.

Siegrist, M., Leins-Hess, R., & Keller, C. (2015). Which front-of-pack nutrition label is the most efficient one, 183-190.

Milo, B., Loreta, C., & Riccardo, P. (2022). Eye-tracking for the study of financial decision-making: A systematic review of the literature. *Journal of Behavioral and Experimental Finance*, 100702.

回应：

感谢您的建议和文献推荐，我们阅读了您推荐的文献并且检索相关领域文献，再次推敲全文用词，不再使用“缺失”、“缺少”、“尚无”等绝对化的词语。由于基金领域的实证证据并不充分，因此我们在文献综述时以复杂决策和风险决策进行类比，而推理出假设。

意见 3：假设 3 的推导依据不足。尤其是“两者关系强度在老手中可能被削弱，而在新手中则不然。”

回应：

我们对引言进行再次推敲，为假设推理提供更有力的证据。在 1.3 中，我们从老手的决策方式和行为表现更为稳定的角度提出假设 3，并补充了相关文献(p.3)。

意见 4：各基金选项/属性的呈现方向模式是否应该匹配？以往研究已经发现因为阅读顺序的影响，横向呈现的刺激属性会被加工更多。

回应：

感谢您提出的这一问题，这是我们之前并未深入思考过的。在您的启发下，我们讨论了信息布局的作用并补充了相关文献(4.1 第二段, p.15)。

经过文献检索，我们认为，虽然我们并未对选项和属性的呈现方向进行匹配，但是这可能并不会对结论造成太大影响。有两方面原因：(1)以往研究也多采用每一行表示一个选项，每一列表示一个属性的设计，这是相对可靠的；(2)投资者在决策时偏向采用基于属性的信息搜索模式，属性加工为竖向加工，这有悖于人们日常的横向阅读习惯，因此本研究的结果不能通过信息的布局进行解释。

意见 5：我对决策质量的定义有存疑，尤其是这个指标是否能定义为“决策质量”。基金 443 实际上是基于非补偿性原则下的最优，因为该基金每个属性都优，但是假如基于补偿性原则而忽略某个属性，比如历史业绩，此时最优的基金就是 102 和 443，而不是只有 443。

回应：

感谢您关于决策质量定义的问题，这也是我们在设计实验材料时着重考虑的内容。我们对您的质疑做如下解释：基金 443 确实是每个属性均最优的选择，这不代表投资者需要通过补偿性原则进行信息加工，研究表明被试在完成类似的信息检索任务时会对信息进行选择 (Lussier & Olshavsky, 1979; Weenig & Maarleveld, 2002)。我们更关注被试在面对诸多信息时究竟会如何检索和加工信息，即会逐一查看所有信息后进行加权比较(补偿性)，还是查看某一属性信息后排除某个选项，随后不再查看该选项信息(非补偿性)。因此，我们关注的补偿性和非补偿性体现在被试的加工策略，而非最优选项。

为了探究个体在最初决定投资基金时的推理过程，我们“需要”在任务中设置最优选项，这个最优选项是在属性间权衡后的结果，反映了逻辑推理后的最优选项。例如，基金 443 和基金 102 年收益率均为 22.36%，但是基金 102 的历史业绩处于劣势，因此权衡之下会排除基金 102，即基金 102 为可接受的选择而非最优选择。我们关注被试能否在初入场时选中逻辑上的最优选择。

意见 6：有些遗憾的是，论文对新手老手在决策过程上的差异，未能做出更清晰的描述。是否有其他指标？建议作者参考如下文献，如果可以深入挖掘老手和新手在基金投资中，在更多信息加工上的本质差异，可能比目前简单的补偿 vs. 非补偿性加工、基于属性/基于选项加工的分析逻辑，提供更多有价值的结果。

Willemsen, M. C., & Johnson, E. J. (2019). (Re) visiting the decision factory: Observing cognition with MouselabWEB. In *A handbook of process tracing methods* (pp. 76-95). Routledge.

回应：

感谢您的意见和文献推荐，这篇文献十分重要，我们之前已经在文中引用，我们再次进行了认真研读，同时也阅读了国内外有关决策过程的实证研究和综述。

在您的建议下，我们阅读了近年来国内有关决策过程的研究，发现它们常以决策时长、补偿性指数和搜索模式作为因变量指标。例如王阿妹等人(2018)同时考虑决策时长、补偿性指数的重要方面——搜索深度，以及搜索模式，探讨个体在复杂决策中的心理过程；黄元娜(2023)以决策反应时和搜索模式作为指标探讨个体在不同决策领域中的决策策略。我们在此基础上，参考 Koele 等人(1995)，同时考虑补偿性指数及其子指标：搜索深度和变异性，试图为理解被试在决策时的心理过程提供更加丰富的证据。

最终，我们分析的过程性指标包括决策时长、搜索深度、搜索变异性、补偿性指数和搜

索模式，这是以往研究常用的指标(例如 Payne, 1976; Reisen et al., 2008; 王阿妹等, 2018)，能够较好地说明被试在决策时的心理过程(见 2.4.1, p.7)。同时，我们在讨论中也对老手和新手的决策过程差异进行了补充。

意见 7: 研究选择的 4 个指标中，从公式和定义看，CI 完全可以替代 DS 和 VS。后二者是否还有存在的必要？

回应:

关于是否同时保留 CI 和 DS、VS 这一问题，我们考虑到以往研究者多使用搜索深度 DS 作为指标，但是 Koele 等人(1995)提出应将搜索深度 DS 和搜索变异性 VS 整合为补偿性指标，作为整体来代表个体在决策时是基于补偿性还是非补偿性。但是补偿性指数通过公式计算得来，理解的难度要高于搜索深度和变异性。所以，我们权衡后决定将这些指标保留，来为补偿性提供更加丰富的证据，也便于与其他使用了其中某一或某些指标的研究对话。

意见 8: 一些涉及研究细节的内容需要补充报告。如：

- 1) 研究 1 请补充报告使用 MouseLab 的版本，并附上参考文献；
- 2) 请补充关于 MouseLab 各指标选择依据、与研究假设关系，及其参考文献。
- 3) 请补充说明研究测量工作记忆的必要性。并报告两组被试在工作记忆上的差异，在后续的统计分析中，是否控制了工作记忆？

回应:

感谢您指出我们在方法写作中的不足。在您的提醒下，我们对研究方法部分进行了更加清晰的说明：在 2.4.1 中补充了 Mouselab 的版本(2.4.1 第一段, p.6)以及各指标选择依据(2.4.1 第四段, p.7)。

同时，在 2.3 中，我们补充了一个自然段用于说明实验流程及各个任务的具体目标(2.3 第二段, p.6)，由于工作记忆是影响老手、新手表现的关键因素，因此我们将其作为控制变量纳入研究。在随后的统计分析中，我们比较了两组被试在工作记忆上的差异(见表 1, p.10)，并将工作记忆作为控制变量纳入回归分析(见表 2, p.12)，还在讨论中补充了为何两组被试在工作记忆存在区别(4.4 第二段, p.17)。

意见 9: 效应量估计不够合理。两个研究均选择了 0.8 的效应量。但选择的依据不合理。在决策/社会心理学领域，0.8 的效应量非常罕见；从两个研究数据看，效应量也仅在 0.50 左右，

0.80 的效应量明显高估了。其中，研究 1 选择了一篇老手、新手在一般决策表现的相关研究 (Sella et al., 2018), Cohen's $d > 1$ 作为依据，但是这篇论文考察了的是“interaction between dyads of individuals with different levels of expertise”，与该论文主题相去甚远。建议选择老手、新手在眼动或其他过程追踪考察多属性决策、投资决策的文献的效应量，作为依据。研究 2 选择樊亚凤等人(2019)研究，主题是默认选项设置对个人捐赠意愿的影响，无论自变量还是因变量都和该论文主题无关，建议检索利用选项结构化表征助推的研究作为效应量的计算依据。建议按照新的效应量重新估计样本量并补充被试。

回应：

感谢您对样本量估算的建议，我们承认本研究样本量的规模不如社会心理学中一般的行为实验那样大，这主要有两方面原因：一为“老手-新手”或“专家-新手”的研究范式，往往意味着找到这些特定的老手和新手的难度比招募普通被试要大很多，故而影响其样本量；二为本研究需要在无人打扰的环境下，逐一让被试完成实验程序，因此实施较为困难，故而难以如集体施测那样使用较大样本。

参考您的建议，我们检索了有关“专家-新手”的决策过程研究，在 2.2(p.6)以及 3.1(p.14)中补充了样本量估算过程。具体而言，两项国内研究可为研究 1 的样本量估算提供参考：(1) 陈梅香和白学军(2019)在《心理科学》上发表的文章，通过眼动技术考察专家医生($n = 15$)和新手医生($n = 19$)的决策过程和正确率，结果发现两组被试在决策正确率($\eta^2 = 0.11$)和决策用时($\eta^2 = 0.28$)上存在差异，均为中等偏大的效果量；(2) 王福兴等人(2016, 研究 1)在《心理学报》上发表的文章，通过眼动技术考察经验棋手($n = 10$)和新手棋手($n = 15$)的决策过程和复盘正确率，结果发现两组被试在复盘正确率($\eta^2 = 0.73$)和决策用时($\eta^2 = 0.11$)上存在差异，也为中等偏大的效果量。可见研究 1 以中等偏大的效果量进行被试量估算是依据的。同样，潘运娴等人(2018, 研究 2)发表在《心理科学》上的文章可为研究 2 样本的估计提供参考，该研究操纵信息呈现方式后，发现凸显信息的条件下被试的信息搜索准确率更高($\eta^2 = 0.21$)，由此研究 2 中 73 名被试也是符合要求的。

意见 10：整体信息搜寻模式的检验，为什么不检验新手老手两组在 SM 值上的差异，同时，需要排除属性/选项呈现方向的影响。

回应：

感谢您对结果和讨论的建议，在原稿的 2.5.2 第一段的末尾，我们已经对两组被试的 SM 值进行差异检验。“在信息搜索模式上，老手组的 SM 值显著低于新手组， $t(72) = -2.20$, $p =$

0.031, Cohen's $d = 0.51$ ”。由此表明与新手相比，老手更偏好基于属性的信息搜索模式。

关于属性/选项的呈现方向，您在前面的建议中也有提及，再次表示感谢，这启发我们讨论了信息布局的作用并补充了相关文献(4.1 第二段, p.15)。具体而言，我们认为选项和属性的呈现方向可能并不会对结论造成太大影响，主要是因为目前的呈现方法较为可靠和常用，同时当前的结果无法用人们日常的横向阅读习惯来解释。

意见 11: 研究 1 中，搜索深度和搜索变异性的结果，老手与新手的差异方向相反，在搜索深度上，老手补偿性更低，但在搜索变异性上，老手的补偿性反而更高。如何解释？

回应:

感谢您的问题，在原稿中由于我们表述不清而造成了误解。实际上，搜索深度和变异性反映了一致的结果，而并非相反。

我们再次对相关措辞进行推敲，以更好地澄清指标的意义：在方法部分中，我们说明了搜索深度越高、搜索变异性越低表明补偿性越高(2.4.1 第四段, p.7)，在结果中老手的搜索深度更低、变异性更高，由此可见老手的补偿性更低(2.5.2, p.12)。

意见 12: “2.5.3 信息搜寻模式对决策质量的影响”中，Logistic 回归，应该将所有控制变量都纳入回归方程。报告回归模型的显著性水平和解释力。建议将组别直接纳入回归，并考察其与各变量的交互。

回应:

我们根据您的意见重新进行了分析并补充了相关内容：我们直接将组别(经验水平)和 SM 值以及两者的交互项作为预测变量纳入回归分析，结果如表 2 所示(p.12)，同时补充了结果的说明。鉴于新手和老手并非如一般实验设计中随机分配形成的两个等同组，他们为具有差异(如控制变量上的差异)的自然分组，可能各自有着不同的变量关系模式，为了更深入地理解信息搜索模式在老手和新手中的作用情况，继续根据经验水平分组进行 Logistic 回归(p.10)，原稿中的 OR 值即为结果的相对效应量指标，在本次修改中，我们又补充了变化的 Nagelkerke R² 的结果。

意见 13: 从表 2 看，工作记忆对决策质量有预测作用，为什么在前面的分析中没有纳入？正确率低于 85%，正是工作记忆偏低的，为什么这些被试被排除在后续分析？

回应:

在您的提醒下，我们分析了新手和老手在工作记忆上的差异(2.5.2, 见表 1, p.10)，并在讨论部分(见 4.4 第二段, p.17)对此差异进行解释。

感谢您关于研究方法的问题，原稿中我们对此任务的说明不够清晰，给您的理解造成了困扰。本研究采用梅高兴等(2021)修订的工作记忆测量任务以及测量标准。具体而言，工作记忆任务包括记忆字母和算式判断这两组子任务，用正确记忆字母的数量作为工作记忆的指标，用算式判断进行记忆干扰，此处正确率 85%是指算式判断的正确率，排除 85%以下的被试数据是为了防止被试只重复记忆字母而不认真进行算式判断，从而导致字母的工作记忆成绩虚高(此时已经不是单纯的工作记忆了，而是包含了长时记忆的可能作用)。为了有助于读者理解，我们在 2.4.2 中进行了补充说明(见 2.4.2, p.8)

意见 14: 研究 2，主要的实验操纵是被试用表格纸(结构化加工组)或白纸(控制组)进行信息整理。但对具体操纵过程说明过于简略，且未进行操纵检验。实验结束后“根据被试的投资选择给予相应的报酬”，如何给予报酬？是否能起到激励的作用？

回应:

感谢您就研究方法撰写提出的建议和问题，我们从研究 2 的操纵方法以及报酬给予两个方面对文章进行修改。

为更好地说明干预的逻辑及方法，我们在研究 1 的讨论部分(2.6, p.11)予以补充：研究目的在于尽可能在低成本条件下改善新手的决策质量，因而具体的操纵过程十分简单(提供表格纸)，我们相信低成本、易操纵的干预手段才更适合推广。其背后的逻辑在于研究 1 发现的矩阵形式有助于基于属性的加工，以及当前理财 APP 将属性置于选项之内不利于基于属性的加工，因此我们通过提供表格纸来引导被试将属性和选项同时进行排列，从而降低基于属性加工的难度。此外，正如您指出的那样，我们未进行操纵检验。为了提高此干预方法的说服力，在 1.4 中，我们补充了表格式的信息呈现为何能够作为干预手段，并且补充了此干预手段的支撑文献(Spence & Brucks, 1997; p.4)。

关于报酬发放，在方法中我们补充“(被试)被告知在投资中的表现将决定着最终获得的实验报酬”(3.3, p.13)，这也是涉及金钱的决策任务中常用的激励被试认真作答的方法。

意见 15: 其他一些需要补充报告的内容，包括：

- 1) 在分析时，各控制变量是否纳入分析或予以控制，需要报告其结果。
- 2) 表 1 中加入统计检验结果，便于阅读。

3) 所有主要测量变量均需要报告描述统计结果，如决策质量，其他控制变量可以根据对因变量是否有影响，选择性报告其描述统计结果。

4) 需要比较新手老手在决策质量上的差异。

5) 回归方程需要报告 R^2 。

回应：

感谢您对结果处理和呈现的建议，我们对此进行一一修改：

(1) 在原稿的基础上(已将性别、年龄、受教育程度、风险承受能力和工作记忆作为控制变量)，又纳入是否财经类专业这一控制变量进行分析(2.5.2 第一段, p.10)，结果详见表 2 (p.12)：

(2) 已在表 1 中补充统计检验结果；

(3) 已在表 1 中增加风险承受能力和工作记忆水平等控制变量的描述性统计结果；同时在 2.5.2 中通过文字陈述对决策质量进行说明(p.10)；

(4) 在 2.5.2 中结果部分，补充了新手和老手的决策质量差异的结果(p.10)；

(5) 原稿中的 OR 值即为回归结果的相对效应量指标，在本次修改中，我们又补充了变化的 Nagelkerke R^2 的结果(p.11)。

意见 16：文中提到“研究 1 与先前类似文献(于泳红, 汪航, 2005)均表明，矩阵式呈现信息时个体多采用基于属性的信息搜索模式”，这恰好说明了研究 1 中信息呈现方式对老手/新手差异的干扰作用，如何排除这一作用，建议在讨论中予以补充。

回应：

您的建议启发我们对矩阵呈现方式做进一步讨论。我们对此内容进行了补充：一方面，我们讨论了属性和选项的横竖方向对结果的影响(4.1 第二段, p.15)，排除了横向阅读习惯的作用；另一方面，我们也讨论了以往矩阵式信息呈现可能简化了现实情境(见 2.6 第二段, p.11)，这一点是很多决策研究都难以避免的，简化的信息呈现可能干扰了决策策略的选择，使得整体补偿性较高。我们在研究局限中也再次补充说明了任务简化现实情境的情况(4.3 第二段, p.17)。

意见 17：“4.2 互联网基金理财的干预措施”，缺少与以往相关助推研究的比较与分析，如以往有关选择架构中改变信息呈现方式的研究，比较其异同及其原因。

回应：

感谢您的建议，这有助于我们挖掘本研究与助推的关系，提高研究的价值。我们检索了相关文献，并在 4.2 第二段(p.16)进一步探讨了当前的干预手段与以往改变信息呈现方式(凸显和移除)的关系和差异。

意见 18:“4.3 研究价值与局限”部分，研究贡献有待改进。如“二是将‘老手(专家)-新手’范式拓展到了基金投资领域”，行为金融领域里专家和新手的差异，是一个研究很多的主题。这一点的贡献并不突出。“三是提出了简单有效的干预措施”，建议明确具体应该如何干预。

研究的不足太过泛泛，建议深入挖掘一下研究的不足之处。

回应:

感谢您指出我们在价值和局限上的不足，我们对此进行了调整。一方面，我们对研究贡献部分进行了补充(见 4.3 第一段, p.17)，从客观需求的角度，陈述了基金投资者人数增长及当前研究鲜有以个体为中心研究的现状，进而提出本研究的研究价值。同时，我们调整了部分措辞，且更加清晰地陈述了干预措施(见 4.3 第二段, p.17)。

另一方面，我们对研究不足进行了补充，主要从研究方法部分(简化现实情境、未考量长期收益)以及被试选取两个方面对本文的不足及未来研究展望进行挖掘(见 4.4, p.17)。

.....

审稿人 3 意见:

本文尝试用 mouselab 技术探讨基金投资中新手和专家的信息加工差异及其与投资结果的关系，并试图干预新手的投资以改进其投资质量，是一个有趣、也有一定应用性的问题。但作为一项严谨的科学论文，作者需要深思、解决以下问题，并补充许多细节，方能使研究成为一个严谨的、无懈可击的、高说服力的故事。

意见 1: 作为一种过程取向的方法，mouselab 技术重在考察多属性决策过程中，决策者的信息偏好或加工方式，较多地使用在候选对象属性多样、相对固定的决策任务(如择偶、择业、购物等)上，而基金、股票等的属性相对较少且较不固定，意味着人们的选择通常可能基于少量属性(如收益率、风险)，使用该技术需要慎重。

回应:

感谢您的提醒，在设计研究时我们比较了相关研究技术后选定了 Mouselab 技术，在您的提醒下我们再次检索相关文献以确保该技术使用的合理性。

对于您提出的问题，我们从 Mouselab 的应用领域及基金决策的特征两方面进行解释：

(1) Mouselab 技术已应用于诸多领域，已有大量研究利用该技术在延迟选择(李晓明, 谢佳, 2012; 李晓明, 蒋松源, 2019)、模型检验(Schulte-Mecklenbeck et al., 2013)、消费决策(李红锋, 2014; 夏小红, 梁宁建, 2010)等领域对信息加工过程展开了探讨。通过 Mouselab 技术可获取的搜索深度、变异性、补偿性指标以及搜索信息转换的搜索模式都是本研究所需要的，也是本研究所关注的主题。(2)与消费、招聘等实际决策场景相似，理财 APP 同样为投资者的决策提供诸多属性信息。正如预研究中，我们摘录了支付宝理财等 APP 的基金页面上的属性信息，大概包括“近一年收益率”、“单位净值”、“费率”、“提涨跌幅”、“风险等级”、“基金行业”、“历史业绩”、“基金评级”、“基金持仓情况”、“基金经理情况”、“基金公司情况”、“基金排名”和“基金规模”等 13 个属性，这些属性都是投资者可能会考虑的内容，因此基金属性并不少也并不固定。

参考文献：

- Schulte-Mecklenbeck, M., Sohn, M., Bellis, E. D., Martin, N., & Hertwig, R. (2013). A lack of appetite for information and computation. Simple heuristics in food choice. *Appetite*, 71(4), 242–251.
- 李红锋. (2014). 时间压力和属性数目对决策中信息加工的影响. *广东技术师范学院学报(自然科学)*, (7), 57–63.
- 李晓明, 蒋松源. (2019). 权力对延迟选择的影响. *心理科学进展*, 27(3), 447–452.
- 李晓明, 谢佳. (2012). 偶然情绪对延迟选择的影响机制. *心理学报*, 44(12), 1641–1650.
- 夏小红, 梁宁建. (2010). 情绪及任务种类对购买决策信息加工的影响. *心理研究*, 3(2), 67–71.

意见 2：本研究探讨基金投资中新手和专家信息加工的差异，究竟更关心基础理论问题，还是现实应用问题？也就是说，人们决策过程是理性的或是有限理性的、专家和新手信息加工差异等问题都是非常普遍而基础的问题，探讨基金投资决策中的此类问题能为上述基础问题贡献新的知识？抑或是已有的发现和知识可以推广应用到基金投资问题上？这涉及本研究的贡献或创新性问题。

回应：

您的问题为我们思考本研究的价值提供了新的思路，经过对研究背景的回顾，我们认为本文更关注现实应用问题。为了更好地凸显我们的研究价值，我们在研究价值部分予以补充(见 4.3 第一段, p.17)。主要从实证研究需要关注基金投资新手的社会现实，以及当前实证研究结果难以为普通投资者所利用的角度入手，提出研究价值是：“本文立足基金投资者个体，挖掘新手和老手决策时的心理过程并提出干预措施，目的并不在于提出基金的函数模型或改进决策理论，而是希望能够引导庞大规模的基金新手正确理财”。

意见 3: 综述中前两个假设都过于浅显而无新意, 后两个的理论推导不充分, 较缺乏说服力, 几个假设之间缺乏丝丝入扣的逻辑递进关系, 整体感觉故事线不明确、不是一个有张力的逻辑体系或故事。

回应:

感谢您指出引言部分的不足, 这启发我们进一步思考研究价值并改善文章逻辑。一方面, 我们对假设 1 和假设 2 的部分措辞进行调整, 以更清楚地体现当前研究与以往研究的差异, 凸显对假设进行验证的必要性(1.1、1.2, p.1-3); 另一方面, 我们对引言进行再次推敲, 补充并调整了假设 3 和假设 4 的理论逻辑: 在 1.3 中, 我们从老手的决策方式和行为表现更为稳定的角度提出假设 3, 并补充了相关文献(p.3); 在 1.4 中, 我们补充了表格式的信息呈现为何能够作为干预手段, 并且补充了此干预手段的文献(Spence & Brucks, 1997; p.4)。

意见 4: 与此相应地, 实验部分较多地说到了做了什么、怎样做的, 较少关于每一个动作或做法理由的交代, 比如为什么选用 mouselab 技术中的那些指标、每个指标旨在解决什么问题、合起来回答什么问题? 给人一种知其然不知其所以然的感觉。

回应:

感谢您指出研究方法撰写的不足, 我们做了如下调整: 对于研究 1, 我们先在 2.3(p.6) 说明了设计任务的原因, 随后在 2.4.1 中(p.6-7)补充了指标选择的标准及原因。对于研究 2, 我们在 2.6(p.11)中补充了干预的逻辑和背景, 为研究 1 和研究 2 之间的逻辑关系提供说明。

意见 5: 结果部分表明专家更基于线索而非补偿, 但新手更受益于基于线索和非补偿, 缺乏必要的理论分析, 给人相互矛盾之感。那么, 专家的加工究竟是好是坏呢? 为什么? 仅仅归结于其一般能力是不够的。

回应:

感谢您的问题, 我们认为结果是不矛盾的, 为了更充分地讨论这一结果, 我们补充阅读了文献, 并对 4.1 第三段进行修改 (p.15)。具体而言, 新手的经验不足, 较好地信息搜索策略和非补偿性原则能够有效地提升他们的决策质量; 但是对于老手来说, 丰富的经验使得他们有着整体较好的信息搜索策略, 并且形成了自己的决策习惯(Gillespie & Peterson, 2009), 此时整体较好的信息搜索策略已经不是影响他们表现的关键原因, 在研究 1 中我们发现工作

记忆水平(Ericsson, 2000)反而更重要。从结果来看,老手的加工策略是更优、更高效的,但是策略仅反映了技术和方法,而更优的方法并不一定导致更优的结果,研究1发现了老手的选择更来源于他们的一般能力而非方法。

意见 6: 研究二的干预研究与研究一发现之间的内在联系过于牵强。如何证明研究二采用的干预方案真的就提高了新手的基于线索的加工了?亦或许仅仅是帮助他们更好地整合了线索而已。此外,还有许多细节需要补充,比如被试数量是如何确定的。

回应:

根据您的建议和问题,我们从干预手段的提出以及被试选择两方面予以调整。

(1) 关于研究1和研究2的关系以及研究2干预手段的选择这一问题,我们在三个地方予以补充说明:分别为引言部分(1.4, p.4)、研究1的讨论部分(2.6, p.11)以及结果的讨论部分(4.2, p.16)。基本逻辑为研究1的结果以及以往决策研究的结果可能反映了当信息得到排列时,个体能够更加直观地获取同一行和同一列的数据,从而更容易关注和识别属性间的差异(Slovic & MacPhillamy, 1974),实际上,很多研究者认为这种矩阵模式呈现信息的方式降低了信息搜索的难度(丁夏齐等, 2004; 余雯等, 2013),但是这种矩阵化呈现却不是现有的基金APP最主要的呈现方式。目前的基金页面中,属性信息往往置于选项之下,投资者要想进行基于属性的搜索难度是较大的,因此,我们参考Spence和Brucks (1997)给决策者提供表格纸的方式,为被试提供结构化引导。这种干预方法暗含了一个逻辑,理财APP原有的呈现方式更加凸显选项,因此让投资者难以进行属性加工,而表格纸帮助他们将选项和属性同时凸显,从而比原有呈现方式更有助于进行属性加工。

(2) 我们检索了“专家-新手”的复杂决策过程研究,在2.2(p.6)以及3.1(p.14)中补充了样本量估算过程。其中研究1主要参考了复杂决策的过程研究中,“专家-新手”比较范式的效果量(陈梅香,白学军, 2019; 王福兴等, 2016);研究2参考了信息呈现方式对信息搜索正确率影响的效果量(潘运娴等, 2018)。

意见 7: 讨论部分未能就本研究的理论和现实价值给出详细有力的挖掘和分析,没有让读者充分感觉到研究的价值和意义。

回应:

在您的建议下,我们从更多的角度对研究进行了讨论,主要从三方面展开:(1)在4.1中,对老手和新手的差异进行了更多的比较,补充了决策质量、决策时长等结果,同时对阅读习

惯可能造成的影响进行了说明(4.1 第二、三、四段, p.15); (2) 在 4.2 中我们补充了矩阵呈现作为助推方法的有效性与应用价值(4.2 第二段, p.16); (3)在 4.3 中我们从客观需求入手,直截了当地对研究意义进行说明,明确提出本文旨在引导庞大规模的基金新手正确理财而非构建基金的函数模型和决策理论(4.3 第一段, p.17)。

第二轮

审稿人 2 意见:

该论文在研究较少的基金投资决策领域,探讨了其信息加工模式,选题较有价值。在第一轮修改中,作者基本回答了我及其他两位审稿人提出的问题。但该论文在几个部分仍存在问题,有待继续修改。

意见 1: 文章对研究贡献的挖掘和强调稍显不足。表现在:

1)行为金融领域对于股票投资决策研究已经非常多,文章提出的基金投资“决策是一种特殊的风险决策”,对于股票投资同样适用。因此本研究提出的问题,新手和老师在信息加工上的差异,对于股票投资来说不是个新问题。建议文章增加一部分综述,总结基金投资决策和股票投资决策有什么特异性的差异,导致难以直接将股票投资领域的研究成果直接拓展到基金投资领域。我个人认为,这是论文的主要贡献之一,应在文章中予以充分补充和强调。

2)作者在对审稿人 1 的意见的回复中提到:“我们致力于探讨投资者进场的问题,即关注个体在最初决定投资基金时的推理过程,在本研究的实验任务中则表现为存在理论上的最优选项的情况下,投资者能否选中最优选项及其背后的认知推理过程。”这一点贡献,应在论文中予以充分强调,包括引言、摘要和讨论部分。

回应:

十分感谢您提出的建议,这对我们突出本研究的贡献有重要意义。在您的启发下,我们细致检索了以往金融决策的相关研究,并强调了本研究的贡献。具体说明如下:

关于基金投资与股票投资的区别,我们进行了如下思考和修改。您提到的新手与老手在股票投资过程中信息加工差异已经比较常见,对此我们再次在 web of science 上进行了文献检索,以 investment experience、financial experience、financial literacy、financial knowledge、financial ability、novice、expert、expertise 为老手-新手变量的检索词,联合 risk decision、venture decision、stock investment、stock decision、fund investment 等词进行检索,共发现 106

篇期刊论文，其中涉及股票投资的共 34 篇。然而这些主要集中于投资者(1)非理性的心理异象(Talwar et al., 2021)、(2)投资风格(Arora & Kumari, 2015)和(3)投资参与度(王佳, 于泳红, 2017)，而很少深入到信息加工过程层面来探讨投资经验与投资决策的信息加工过程的关系这一问题。对此，我们有理由相信新手-老手的信息加工差异不仅在基金决策领域是一个较新的问题，也对包括股票投资在内的金融决策领域有所启示。为此，我们在正文中予以相应内容的陈述(1.1, p.1 第一段)。

此外，在您的建议下，我们对摘要和引言进行了调整和补充，强调本研究致力于探讨“投资者进场的问题”这一贡献(p.1 摘要； p.1 引言)。

意见 2：研究提出的假设存在部分不足。

1)假设 3 的推理存在部分矛盾：“假设 2：相比于新手，老手在基金投资决策中……更遵循基于属性的搜索模式”；“假设 3：基于属性的信息搜索模式带来的决策质量更好，……两者关系的强度在老手中可能被削弱”。为什么对于基于属性搜索的老手，决策质量更好的效应会被削弱？这里缺乏令人信服的理论推导依据。按照前文提出的“专家的表现关键在于工作记忆”、“老手的表现优劣不在于策略，而是与一些能力有关”，似乎更合理的推论是老手的决策质量更高？

2)假设 4 需要明确表述。“通过引导被试对信息进行属性的结构化整合能够提升基金投资决策的质量。”这里的“被试”效应是对所有人，还是只对新手？

回应：

感谢您指出的问题。假设 3 提出信息搜索模式的作用在老手中被削弱主要基于老手信息加工较为稳定且已经形成自己的投资风格，我们并未对老手和新手的决策质量差异进行假设，因此从假设来说并不矛盾。

而针对老手虽更基于属性进行加工，但其决策质量并不高这一结果，我们认为这主要是因为老手的决策质量与属性加工模式无关。我们从两个方面对讨论进行补充。第一，老手已经更多地且更稳定地依赖属性搜索模式来决策，他们在信息搜索模式上的低变异性可能导致信息搜索模式无法在实际决策中或统计结果上为决策质量提供帮助和解释。第二，老手的决策质量更受制于一般认知能力或过度自信等个体特征(4.1, p.16, 第三、四段)。同时，我们在 4.4 局限及展望中(p.19 最后一段)补充强调了老手的信息搜索模式更为稳定这一结果。

我们调整了假设 4 的措辞，将干预的对象界定为新手并补充了原因(1.4, p.4)。

意见 3: 在样本量估计上, 存在明显的错误。研究 1 中将按照 Cohen's $d = 0.7$ 估计样本量, 但是 power 按照 0.8 计算? 如果是按照 power = 0.7 估计, 独立样本 t 检验的样本量应该为每组 45 人。研究 2 中参考的 ϕ 值是 .52, 但在计算的时候 power 还是 0.8? 即使是独立样本 t 检验, 在 power = 0.8 的时候, 样本量也是每组 35, 而不是 31。建议作者根据修改后提出的效应量及估计出的正确样本量, 补足两个研究的被试数。此外, “论文自检报告”中效应量估计部分内容, 需要根据已经进行的修改进行更新。

回应:

感谢您提出的问题, 我们再次按要求核对了样本量。同时, 本研究效果量(Cohen's d 和 ϕ) 的大小是基于以往的类似研究选择的, 而统计检验力 power(1- β)则是采用了常用的 0.8 作为指标, 两者的内涵并不相同。在以往研究中, 两者的数值也并不要求一致。感谢您的提醒, 我们对自检报告的相应内容进行了修改。

意见 4: 论文中存在部分干扰变量, 其作用未被充分排除。

1)对老手的操作定义可能对研究结果有一定污染。这主要表现在老手和新手在决策质量上的差异。论文现在定义的老手以投资经验的时长为主要指标, 但是投资时间长不一定必然投资收益更高。从以往报告看(景顺长城等, 2021), 基金赚钱基民“不怎么赚钱”, 50%以上的基民投资收益率低于 0%。因此, 很可能老手中的一部分人投资决策的经验丰富, 形成了稳定的信息加工模式, 但是其投资质量并不高。这可能是研究 1 老手新手决策质量没有显著差异的原因之一。建议文章对此进行充分讨论和分析。

2)在分析的鼠标追踪指标上, 现在以信息搜索相关指标为主。但是信息搜索是以信息已经被充分加工为基础的, 如果各基金的信息并未得到充分加工, 基于不完整信息的信息搜索和基于完整信息的搜索, 可能存在较大差异。因此, 建议补充或控制决策前点击信息的数量的相关指标。

回应:

十分感谢提出的建议, 这促使我们进一步思考我们的研究结果。我们认为老手决策质量并不高于新手的这一结果是因为老手的决策质量与属性加工模式无关, 因此较高效的信息搜索模式不能为老手决策提供助益。一方面, 老手已经更多地且更稳定地依赖属性搜索模式来决策, 他们在信息搜索模式上的低变异性可能导致信息搜索模式无法在实际决策中或统计结果上为决策质量提供帮助和解释。另一方面, 老手的决策质量更受制于一般认知能力或过度自信等个体特征(4.1, p.16, 第三、四段)。

此外，您提到应控制点击信息的数量，我们对此予以尝试。结果显示：在其他条件不变的情况下，新增加工单元格数作为控制变量后，主要结果的方向和显著性保持不变，仍支持假设 3。我们更新了研究 1 的结果部分的文字部分及表格中的所有数据(2.5.3, p.11 第一段和 第二段；表 2, p.13)。

意见 5：论文还有一些小问题有待修改：

1)对部分变量需要有更明确的定义，并报告其描述统计结果。如 2.5 中，应该报告各分析变量的描述统计结果。提到，“其信息加工的时间、信息搜索模式数据以及最终决策将被记录，这些指标均能反映被试在决策时的心理过程”，这里具体指什么指标？各对应什么心理过程？表 2 中建议修改，只报告标准化的回归系数。并在回归表中附上模型估计的结果。

2)研究 2 中，是否进行了操纵检验，检验干预组被试是否真的进行了结构化排列？

3)1.1 部分第一句，“在以往基金投资的研究中，研究者多采用基于结果的研究范式，通过构建数据模型对过程进行推测，但多种模型间存在竞争且难以证伪(Johnson et al., 2008; 魏子晗, 李兴珊, 2015)。”这句话的“在以往基金投资的研究中”不准确。两篇文献都说的是风险决策的研究。

回应：

十分感谢您提出的建议和问题。针对您提到的第一点，我们在 2.4.1(p.8)补充说明了这些指标反映了何种心理过程并在表 1 中汇报描述性统计结果。我们斟酌后仍汇报了非标准化的结果。我们出于以下考虑：参考美国心理学会发布的准则建议“纯粹应用研究汇报非标准化回归系数，纯粹理论性研究使用标准化回归系数，当研究同时具备上述两种情况时，两种系数都可以用”。本研究更偏向于应用研究。同时，现在常见的统计分析软件多无法直接呈现逻辑回归的标准化系数，虽可进行人工计算，但是标准化结果在以往逻辑回归的研究中仍较少汇报，例如 Smith 等人(2023)发表在 *Journal of Personality and Social Psychology* 上的文章以及应梦婷等人(2016)发表在心理学报上的文章均使用了 *B* 和标准误的形式汇报结果。综合应用范围、计算难度和现在常用结果呈现形式三方面原因，我们汇报了非标准化结果。

其次，我们未进行操纵检验，但是我们推测这并不会造成问题，主要有两方面原因：一方面，为了提高此干预方法的说服力，在第一轮的修改中，我们补充了表格式的信息呈现为何能够作为干预手段，并且补充了此干预手段的支撑文献(Spence & Brucks, 1997; 1.4, p.4)。另一方面，以往的操纵检验多适用于对特别心理过程或状态的“间接”操纵中，由于目标心理变量不是直接改变的，故需要检验操纵效果。然而，我们是在研究一种客观的操纵如何影

响，并非是其启动的心理状态或认知结构如何影响行为。这种客观操纵是所见即所得的，再做操纵检验反而容易产生无关因素。因而，以往的助推干预的研究往往是不涉及操纵检验的(例如 Pechey et al., 2013; 何铨, 江程铭, 2020), 我们也希望通过提供少而简的信息以改变人们的行为。因此我们在设计实验时并未增加操纵检验的相关内容。

最后，在您的提醒下，我们修改了 1.1 第一句话的表述(p.1)。

.....

审稿人 4 意见:

作者通过“专家-新手”或者“老手-新手”比较范式揭示其基金投资决策结果以及背后信息加工方式的差异，为增进新手的投资决策能力提供干预思路。是一篇比较有现实意义的研究。

意见 1: 作者在摘要中提到：为了引导众多的基金投资新手正确理财，有必要采用“老手-新手”比较范式探寻二者投资决策时的信息加工过程差异，该问题以往很少被关注。这显然有失偏颇，关于消费者知识有许多相关研究，建议作者修改。

回应:

感谢您的建议。我们细致检索了以往金融决策的相关研究，并强调了本研究的贡献。我们再次在 web of science 上进行了文献检索，以 investment experience、financial experience、financial literacy、financial knowledge、financial ability、novice、expert、expertise 为老手-新手变量的检索词，联合 risk decision、venture decision、stock investment、stock decision、fund investment 等词进行检索，共发现 106 篇期刊论文。然而这些主要集中于投资者(1)非理性的心理异象(Talwar et al., 2021)、(2)投资风格(Arora & Kumari, 2015)和(3)投资参与度(王佳, 于泳红, 2017), 而很少深入到信息加工过程层面来探讨投资经验与投资决策的信息加工过程的关系这一问题。对此，我们有理由相信新手-老手的信息加工差异不仅在基金决策领域是一个较新的问题，也对包括股票投资在内的金融决策领域有所启示。为此，我们在正文中予以相应内容的陈述(1.1, p.1 第一段)。

意见 2: 作者的贡献到底在理论层面还是在实践层面？回顾以往研究，包括作者实验操作，更像是一个有较强实践意义的研究。如果是这样，建议作者重新思考引言部分，目前的写作很难有说服力。

回应:

感谢您指出的问题，我们再次对本文的研究意义进行思考。综合来看，本研究的贡献在于理论和实践两个层面。

从理论上讲，过往研究多采用基于结果的研究范式，多聚焦投资者非理性的心理异象、投资风格或投资参与度(Arora & Kumari, 2015; Talwar et al., 2021; 王佳, 于泳红, 2017)，较少直接刻画投资决策过程。因此，本研究是基于以往实证资料的不足而开展的，为此我们在1.1中予以补充(p.1 第一段)。同时，我们将在医疗等风险决策中常见的“专家-新手”或者“老手-新手”比较范式推广至基金投资领域来刻画基金投资者在决策时的信息加工差异(1.2 老手、新手的加工差异)。即本研究揭示了老手、新手信息投资决策的过程特点，也注重将过程研究范式应用于基金投资这一相对较新的领域。

从实践来讲，本研究为新手的基金决策提供指导。为此，我们强调了以往的复杂模型虽具有科学性但难以被普通基金投资者所理解和应用(引言第一段, p.1)，并突出我们的目的“致力于探讨投资者进场的问题” (1.5, p.5)。

意见 3: 作者提出的假设 1 是否有必要?

回应:

我们认为对假设 1 进行检验是有必要的，在本次修改中我们补充说明了“国内外目前对基金投资决策的研究还较少，包括基金、股票等各类投资决策的相关研究也多关注投资者非理性的心理异象、投资风格或投资参与度，较少直接探讨其背后的信息加工过程。” (1.1, p.1 第一段)，虽然通过消费和赌博任务的研究推测被试的偏好，但是缺少投资领域的直接证据，因此我们予以检验。

意见 4: 作者提到老手更遵循基于属性的搜索模式(假设 2)，然后在假设 3 中又提出基于属性的信息搜索模式带来的决策质量更好，与新手投资者相比，两者关系的强度在老手中可能被削弱。两者是否矛盾?

回应:

感谢您提出的问题。假设 2 提出老手更遵守基于属性的信息搜索模式，假设 3 提出信息搜索模式的作用在老手中被削弱，这主要基于老手信息加工较为稳定且已经形成自己的投资风格，我们并未对老手和新手的决策质量差异进行假设，因此从假设来说并不矛盾。

而针对老手虽更基于属性进行加工，但其决策质量并不高这一结果，我们认为这主要是因为老手的决策质量与属性加工模式无关。对此，我们从两个方面对在讨论中进行补充。第

一，老手已经更多地且更稳定地依赖属性搜索模式来决策，他们在信息搜索模式上的低变异性可能导致信息搜索模式无法在实际决策中或统计结果上为决策质量提供帮助和解释。第二，老手的决策质量更受制于一般认知能力或过度自信等个体特征(4.1, p.16, 第三、四段)。同时，我们在 4.4 局限及展望中(p.19 最后一段)补充强调了老手的信息搜索模式更为稳定这一结果。

意见 5: 决策质量这一构念的界定与测量，其科学依据是什么？

意见 6: 根据上一轮的修改，作者添加了更多细节的描述，基金 443 也为最优选项，但被试最终选中该最优选项，则决策质量记为 1，若未选中则记为 0。这样的做法是否恰当？其他几支基金是否有相对的优劣之分？

回应:

第 5 和第 6 个问题均关注决策质量的界定和操作。关于决策质量的测量，我们承认若能直接调查投资者过往的收益率则更为可靠，但是这一指标无法在新手中进行测量，因为他们几乎没有投资经历。同时，若强制要求被试参与真实的基金投资项目并记录其未来的收益和损失情况，又会引发伦理问题。

因此我们只能选择在实验室中进行模拟，那必然要设置最优选项。设置最优选项是为了提供逻辑推理的标准，其他选择并未设置相对的优劣之分。此次修改中，我们在 1.5 本研究目的和方法的第一段(p.5)对决策质量的设置进行了再次说明。

关于以往模拟决策质量的研究，也往往人为设置标准。例如，Ognjanovic 等人(2019)调查股票投资新手和老手的差异，以他们对几项股票风险感知判断的矛盾性作为决策质量的指标。基于此，我们也选择了重要性相对较高的 5 个基金属性，人为设置属性的表现，以模拟基金投资情境。

意见 7: 作者参考以往研究设计了 6(选项) × 5(属性)的信息板，如果选项和属性变少，或者变多，作者所提出的效应会受到哪些影响？建议添加到讨论里。

回应:

感谢您提出的建议，这启发我们丰富研究讨论并思考现实意义。我们认为与老手相比，新手更容易受到信息的信息复杂度的影响，即在真实环境中新手将更容易犯错，基于属性的信息搜索策略将对新手来说更为有效。我们补充了相关讨论(4.2, p.18)。

意见 8: 作者为什么没有采用操纵的方式来操纵被试的主观知识(即新手 vs. 老手)?

回应:

关于为何没有操纵被试主观知识的问题,与您探讨。对被试的主观心理因素进行操纵是社会心理学常用的方法,似乎能够解释知识与决策的因果关系,但这与我们的研究目的无关。我们希望回答的问题是“如何改善基金投资新手的决策质量”而非“是否需要丰富基金投资新手的理财经验”,若对主观知识进行操纵似乎引发的现实思考是应推行财经教育或干预主观判断,“但是这种财经教育往往需要社会的共同配合,工程量大且见效缓慢”(1.4, p.4 第一段),这与我们“定位在基民本身”方便而快捷的改善其决策质量的初衷相背离。

第三轮

审稿人 2 意见: 作者已经尽力回复了我的问题。虽然对于效应量等问题,我并不感觉很满意。但是这些问题在作者不增加样本的前提下无法解决。因此留给编委做出最终决定。

审稿人 4 意见: 作者很好地回答了我的疑问,建议接收。英文摘要建议找专业人员润色一下,目前翻译痕迹比较重。

回应: 谢谢您的建议,我们已请专业人员对英文摘要进行润色,润色后的英文摘要见 p.27-p.28。

第四轮

编委意见:

本文的研究问题比较新颖,也比较有现实意义。经过几轮的修改,文章在文献综述,严谨性和描述的准确度有了很大的进步。建议接收。

主编意见:

文章有创新性贡献,研究方法有新意。仅有研究的统计功效问题与作者商榷:

研究 1:

计划：参考国内“老手-新手”决策过程研究的结果(陈梅香, 白学军, 2019), 选择中等偏上的效果量 Cohen's $d = 0.7$ 估算样本量, 在 power 为 0.8, $\alpha = 0.05$ 的情况下, G*Power 计算样本量为 68 人, 【按 Cohen's $d = 0.50$ 计划样本量为 128 人】

实际：在补偿性各指标上, 老手组的搜索深度显著低于新手组, $t(72) = -2.20$, $p = 0.031$, Cohen's $d = 0.51$; 老手组的搜索变异性略高于新手组, 两组间差异边缘显著, $t(72) = 1.95$, $p = 0.056$, Cohen's $d = 0.47$; 老手组的补偿性指数显著低于新手组, $t(72) = -2.15$, $p = 0.035$, Cohen's $d = 0.50$, 效果量为中等水平。

研究 2:

计划：参考信息呈现方式影响信息搜索正确率的效果量($\phi = 0.52$; 潘运娴等, 2018), 在 power 为 0.8, $\alpha = 0.05$ 的情况下, G*Power 计算样本量为 31 人, 本研究样本量符合要求。

【按 $\phi = 0.24$ 计划样本量为 90 人】

实际：结果表明结构化加工组的决策质量(正确人数 17 人, 正确率 47.22%)显著高于控制组(正确人数 9 人, 正确率 24.32%), $\chi^2 = 4.17$, $p = 0.041$, $\phi = 0.24$, 效果量为中等水平, 建议作者补充被试或把 inadequate power 作为重要局限讨论。

回应:

感谢您的建议。我们在 4.4 局限及展望部分(p.19, 第一段)补充了关于本研究统计检验力的讨论。总的来说, 我们认为虽然本研究的实际效果量并未达到预估样本量时使用的效果量水平, 但这并不影响研究的统计效度。具体而言, 较小的样本量可能导致统计检验力不足, II 类错误风险增大, 即本来应该接受备择假设时却由于结果不显著而接受了虚无假设, 出现了虚假的阴性结果。然而整体而言, 本研究拒斥了虚无假设, 所以不存在类似问题。在您的提醒下, 我们在正文中提醒未来研究者进行类似实验时, 应选择更为保守的效果量进行样本量估计, 以免出现结果不显著的虚假阴性结果。

第五轮

主编意见:

关于作者的回复, 我有一些个人看法与作者分享:

很多文献论述过被试量不足的低效力的研究不仅容易犯 II 类错误, 而且 I 类错误的风险也可能增加, 导致不精确的估计, 和可重复性的降低。如以下这篇:

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., ... &

Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European journal of personality*, 27(2), 108–119.

“Among other factors, publishing many low-powered studies contributes to this excessive false-positive bias. It cannot be stressed enough that researchers should collect bigger sample sizes, and editors, reviewers, and readers should insist on them.”

“A study with low power is, by definition, unlikely to obtain a significant result with a given effect size. Unlikely events sometimes happen, and underpowered studies may occasionally obtain significant results. But a series of such results begins to strain credulity. In fact, a series of under-powered studies with the same result. ... Such problems are very common in many recently published studies.”

所以还是建议作者补充被试，或把 *inadequate power* 作为重要局限加强讨论。

回应：

感谢您的文献推荐，我们在此基础上进行了思考和补充，同时也在文中引用了这篇文献。在 4.4 局限及展望(p.18)中，我们将统计检验力不足放在第一点局限中进行讨论，以凸显对这点局限的重视，我们说明了样本量不足可能导致的统计效度风险，并解释了我们样本量有限的原因。

第六轮

主编意见：同意发表。